(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 612 033 B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
24.11.1999 Bulletin 1999/47

(51) Int Cl.[6]: G06F 15/80

(21) Application number: 94107842.0

(22) Date of filing: 15.09.1989

(54) **learning processing system**

Lernverarbeitungssystem

Système de traitement d'apprentissage

(84) Designated Contracting States:
DE FR GB

(30) Priority: 17.09.1988 JP 23284588
17.09.1988 JP 23284688
20.09.1988 JP 23544188

(43) Date of publication of application:
24.08.1994 Bulletin 1994/34

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
89402544.4 / 0 360 674

(73) Proprietor: SONY CORPORATION
Tokyo (JP)

(72) Inventor: Hiraiwa, Atsunobu
Shinagawa-ku, Tokyo (JP)

(74) Representative: Thévenet, Jean-Bruno et al
Cabinet Beau de Loménie
158, rue de l'Université
75340 Paris Cédex 07 (FR)

(56) References cited:
- IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, vol.1, 24 July 1988, SAN DIEGO, USA pages 363 - 370 KUNG 'An algebraic projection analysis for optimal hidden units size and learning rates in back-propagation learning'
- IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, vol.1, 24 July 1988, SAN DIEGO, USA pages 247 - 254 KOUTSOUGERAS 'Training of a neural network for pattern classification based on an entropy measure'
- NEURAL NETWORKS FROM MODELS TO APPLICATIONS, PROCEEDINGS OF THE NEURO CONFERENCE, 6 June 1988, PARIS, FRANCE pages 178 - 186 EKEBERG 'automatic generation of internal representations in a probabilistic artificial neural network'
- NEURAL NETWORKS FROM MODELS TO APPLICATIONS, PROCEEDINGS OF THE NEURO CONFERENCE, 6 June 1988, PARIS, FRANCE pages 663 - 672 COCHET 'zzenn zig zag epigenetic neural networks and their use in information systems'
- IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, vol.1, 24 July 1988, SAN DIEGO, USA pages 325 - 333 SIETSMA 'Neural net pruning - why and how'

EP 0 612 033 B1

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

Field of the invention

[0001] This invention relates to a signal processing apparatus or system carrying out signal processing with the use of a so-called neural network made up of a plurality of units each taking charge of signal processing corresponding to that of a neuron, and a learning processing apparatus or system causing a signal processing section by said neural network to undergo a learning processing in accordance with the learning rule of back propagation.

Prior art

[0002] The learning rule of back propagation, which is a learning algorithm of the neural network, has been tentatively applied to signal processing, including high speed image processing or pattern recognition, as disclosed in "Parallel Distributed Processing", vol. 1, The MIT Press, 1986 or "Nikkei Electronics", issue of August 10, 1987, N° 427, pp 115 to 124. The learning rule of back propagation is also applied, as shown in Fig. 1, to a multistorey neural network having an intermediate layer 2 between an input layer 1 and an output layer 3.

[0003] Each unit u_j of the neural network shown in Fig. 1 issues an output value which is the total sum net_j of output values O_i of a unit u_i coupled to the unit u_j by a coupling coefficient W_{ji}, transformed by a predetermined function f, such as a sigmoid function. That is, when the value of a pattern p is supplied as an input value to each unit u_j of the input layer 1, an output value O_{pj} of each unit u_j of the intermediate layer 2 and the output layer 3 is expressed by the following formula (1)

$$O_{pj} = f_j(net_{pj}) = f_j(\Sigma_i W_{ji} \cdot O_{pji}) \qquad (1)$$

[0004] The output value O_{pj} of the unit u_j of the output layer 3 may be obtained by sequentially computing the output values of the inputs u_j, each corresponding to a neuron, from the input layer 1 towards the output layer 3.

[0005] In accordance with the back-propagation learning algorithm, the processing of learning consisting in modifying the coupling coefficient W_{ji} so as to minimize the total sum E_p of square errors between the actual output value O_{pj} of each unit u_j of the output layer 3 on application of the pattern p and the desirable output value t_{pj}, that is the teacher signal,

$$E_p = \frac{1}{2} \sum_i (t_{pj} - O_{pj})^2 \qquad (2)$$

is sequentially performed from the output layer 3 towards the input layer 1. By such processing of learning, the output value O_{pj} closest to the value t_{pj} of the teacher signal is output from the unit u_j of the output layer 3.

[0006] If the variant ΔW_{ji} of the coupling coefficient W_{ji} which minimizes the total sum E_p of the square errors is set so that

$$\Delta W_{ji} \ \alpha - \partial Ep/\partial W_{ji} \qquad (3)$$

the formula (3) may be rewritten to

$$\Delta Wji = \eta \cdot \delta_{pj} O_{pj} \qquad (4)$$

as explained in detail in the above reference materials.

[0007] In the above formula (4), η stands for the rate of learning, which is a constant, and which may be empirically determined from the number of the units or layers or from the input or output values. δ_{pj} stands for the error proper to the unit u_j.

[0008] Therefore, in determining the above variant ΔW_{ji}, it suffices to compute the error δ_{pj} in the reverse direction, or from the output layer towards the input layer of the network.

[0009] The error δ_{pj} of the unit u_j of the output layer 1 is given by the formula (5)

$$\delta_{pj} = (t_{pj} - O_{pj})f'_j(net_j) \tag{5}$$

On the other hand, the error δ_{pj} of the unit u_j of the intermediate layer 2 may be computed by a recurrent function of the following formula (6)

$$\delta\ pj = f'_j(net_j) \sum_k pk\ W_{kj} \tag{6}$$

using the error δ_{pk} and the coupling coefficient W_{kj} of each unit u_k coupled to the unit u_j, herein each unit of the output layer 3. The process of finding the above formulas (5) and (6) is explained in detail in the above reference materials.

[0010] In the above formulas, $f'_j(net_j)$ stands for the differentiation of the output function $f_j(net_j)$.

[0011] Although the variant W_{ji} may be found from the above formula (4), using the results of the formulas (5) and (6), more stable results may be obtained by finding it from the following formula (7)

$$\Delta\ W_{ji}(n+1) = \eta . \delta_{pj} O_{pi} + \alpha . \Delta\ W_{ji(n)} \tag{7}$$

with the use of the results of the preceding learning. In the above formula, α stands for a stabilization factor for reducing the error oscillations and accelerating the convergence thereof.

[0012] The above described learning is repeated until it is terminated at the time point when the total sum E_p of the square errors between the output value O_{pj} and the teacher signal t_{pj} becomes sufficiently small.

[0013] It is noted that, in the conventional signal processing system in which the aforementioned back-propagation learning rule is applied to the neural network, the learning constant is empirically determined from the numbers of the layers and the units corresponding to neurons or the input and output values, and the learning is carried out at the constant learning rate using the above formula (7). Thus the number of times of repetition $\underline{n}$ of the learning until the total sum E_p between the output value O_{pj} and the teacher signal t_{pj} becomes small enough to terminate the learning may be enormous to render the efficient learning unfeasible.

[0014] Also, the above described signal processing system is constructed as a network consisting only of feedforward couplings between the units corresponding to the neurons, so that, when the features of the input signal pattern are to be extracted by learning the coupling state of the above mentioned network from the input signals and the teacher signal, it is difficult to extract the sequential time series pattern or chronological pattern of the audio signals fluctuating on the time axis.

[0015] In addition, while the processing of learning of the above described multistorey neural network in accordance with the back-propagation learning rule has a promisingly high functional ability, it may occur frequently that an optimum global minimum is not reached, but only a local minimum is reached, in the course of the learning process, such that the total sum E_p of the square errors cannot be reduced sufficiently.

[0016] Conventionally, when such local minimum is reached, the initial value or the learning rate η is changed and the processing of learning is repeated until finding the optimum global minimum. This results in considerable fluctuations and protractions of the learning processing time.

[0017] The paper by Kung et al entitled "An Algebraic Projection Analysis for Optimal Hidden Units Size and Learning Rates in Back-Propagation Learning" (IEEE International Conference on Neural Networks, San Diego, California, July 24-27, 1988) seeks to optimize a learning process consisting of iterative application of the back propagation algorithm and uses an approach whereby it is sought to minimise the value of a modified measure of mean squared error.

[0018] The paper by Koutsougeras et al entitled "Training of a Neural Network for Pattern Classification Based on an Entropy Measure" (IEEE International Conference on Neural Networks, San Diego, California, July 24-27, 1988) discusses a neural network having a branching structure designed to partition n-dimensional space into different discrete regions corresponding to respective different classes of patterns. The branching structure is incrementally built up during a training phase where input patterns are applied to the individual neurons of the network and threshold and weight values of the neurons are adjusted starting from the input layer and proceeding towards the output laser. Extra neurons are added as needed in order adequately to partition the n-dimensional space.

Objects of the invention

[0019] It is a primary object of the present invention to provide a learning processing system in which the signal processing section of the neural network is subjected to learning processing in accordance with a back-propagation learning rule, wherein the local minimum state in the learning processing process may be efficiently avoided for realizing an optimum global minimum state quickly and stably.

Summary of the invention

[0020] For accomplishing the primary object of the present invention, the present invention provides a learning processing system in which the learning processing section executes the learning processing of the coupling strength coefficient as it increases the number of units of the intermediate layer.

[0021] The above and other objects and novel features of the present invention will become apparent from the following detailed description of the invention which is made in conjunction with the accompanying drawings and the new matter pointed out in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0022] Fig. 1 is a diagrammatic view showing the general construction of a neural network to which the back-propagation learning rule is applied.

[0023] Fig. 2 is a block diagram schematically showing the construction of an illustrative example of a signal processing system.

[0024] Fig. 3 is a diagrammatic view of a neural network showing the construction of the signal processing section of the signal processing system according to the system shown in Fig. 2.

[0025] Fig. 4 is a flow chart showing the process of learning processing in the learning processing section constituting the signal processing system of the system shown in Fig. 2.

[0026] Fig. 5 is a block diagram schematically showing the construction of the learning processing system according to the present invention.

[0027] Figs. 6A and 6B are diagrammatic views showing the state of the signal processing section at the start and in the course of learning processing in the learning processing system shown in Fig. 5.

[0028] Fig. 7 is a flow chart showing a typical process of learning processing in the learning processing section constituting the learning processing system shown in Fig. 5.

[0029] Fig. 8 is a chart showing the typical results of tests of learning processing on the signal processing section of the neural network shown in Fig. 5 by the learning processing section of the learning processing system.

[0030] Fig. 9 is a chart showing the results of tests of learning on the signal processing section of the neural network shown in Fig. 3, with the number of units of the intermediate layer fixed at six.

[0031] Fig. 10 is a chart showing the results of tests of learning on the signal processing system of the neural network shown in Fig. 3, with the number of units of the intermediate layer fixed at three.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0032] By referring to the drawings, certain preferred embodiments of the present invention will be explained in more detail.

[0033] An illustrative example of signal processing system will be hereinafter explained.

[0034] As shown schematically in Fig. 2, the signal processing system of the present illustrative example includes a signal processing section 30 for obtaining the output value O_{pj} from the input signal patterns p and a learning processing section 40 for causing the signal processing section 30 to undergo learning to obtain the output value O_{pj} closest to the desired output value t_{pj} from the input signal patterns p.

[0035] The signal processing section 30 is formed, as shown in Fig. 3, by a neural network of a three-layer structure including at least an input layer L_I, an intermediate layer L_H and an output layer L_O. These layers L_I, L_H and L_O are constituted by units u_{I1} to u_{Ix}, u_{H1} to u_{Hy}, and u_{O1} to u_{Oz}, each corresponding to a neuron, respectively, where x, y and z stand for arbitrary members. Each of the units u_{H1} to n_{Hy} and u_{O1} to u_{Oz} of the intermediate layer L_H and the output layer L_O is provided with delay means and forms a recurrent network including a loop LP having its output $O_{j(t)}$ as its own input by way of the delay means and a feedback FB having its output $O_{j(t)}$ as an input to another unit.

[0036] In the signal processing system 30, with the input signal patterns p entered into each of the units u_{I1} to u_{Ix} of the input layer L_I, the total sum netj of the inputs to the units u_{H1} to u_{Hy} of the intermediate layer L_H is given by the following formula (17):

$$\mathrm{net}_j = \sum_{e}^{x} \sum_{k=0}^{NI} w_{jx*k+e}\, O_{ie}(t-k)$$
$$+ \sum_{e}^{y} \sum_{k=1}^{NH} w_{jy*k+i}\, O_{hi}(t-k)$$
$$+ \sum_{1}^{z} \sum_{k=i}^{N0} w_{jz*k+i}\, O_{oi}(t-k)$$
$$+ \theta j \qquad (17)$$

Each of the units u_{H1} to u_{Hy} of the intermediate layer L_H issues, for the total sum net_j of the input signals, an output value $O_{Hj(t)}$ represented by the sigmoid function of the following formula (18):

$$O_{Hj(t)} = \frac{1}{1+e^{-netj}} \qquad (18)$$

[0037] The total sum net_j of the inputs to the units u_{O1} to u_{Oz} of the output layer L_O is given by the following formula (19):

$$\mathrm{net}_j = \sum_{1}^{x} \sum_{k=0}^{NH} w_{jx*k+i}\, O_{Hi(t-k)}$$
$$+ \sum_{1}^{z} \sum_{k=1}^{NO} w_{jz*k+i}\, O_{Hi(t-k)} \qquad (19)$$
$$+ \theta_j$$

While each of the units u_{O1} to u_{Oz} of the output layer L_O issues, for the total sum net_j of the inputs, an output value $O_{oj(t)}$ represented by the following formula (20):

$$O_{oj}(t) = \frac{1}{1+e^{-netj}} \qquad (20)$$

where O_j stands for a threshold value and NI, NH and NO stand for the numbers of the delay means provided in the layers L_I, L_H and L_O, respectively.

[0038] The learning processing section 40 computes the coefficient W_{ji} of coupling strength between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix}, from the output layer L_O towards the input layer L_I, sequentially and repeatedly, according to the sequence shown in the flow chart of Fig. 4, while executing the learning processing of the coupling coefficient W_{ji} so that the total sum of the square errors LMS between the desired output value t_{pj} afforded as the teacher signal and the output value O_{oj} of the output layer L_O will be sufficiently small. By such learning processing, the learning processing section 40 causes the output value O_{oj} of the output layer L_O to be closest to the desired output value t_{zr}, afforded as the teacher signal patterns, for an input signal pattern $p_{(xr)}$ supplied to the signal processing section 30. This pattern $p_{(xr)}$ represents an information unit a whole which fluctuates along the time axis and represented by the xr number of data, where r stands for the number of times of sampling of the information unit and x the number date in each sample.

[0039] That is, the section 40 affords at step 1 the input signal patterns $p_{(xr)}$ to each of the units u_{I1} to u_{Ix} of the input layer L_I, and proceeds to computing at step 2 each output value $O_{pj(t)}$ of each of the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} of the intermediate layer L_H and the output layer L_O.

[0040] The section 40 then proceeds to computing at step 3 the error δ_{pj} of each of the units u_{O1} to u_{Oz} and u_{H1} to u_{Hy}, from the output layer L_O towards the input layer L_I, on the basis of the output values $O_{pj(t)}$ and the desired output value t_{zr} afforded as the teacher signal.

[0041] In the computing step 3, the error δ_{oj} of each of the units u_{O1} to u_{Oz} of the output layer L_O is given by the following formula (21):

$$\delta_{oj} = (t_{pj} - O_{oj})O_{oj}(1 - O_{oj}) \quad (21)$$

wherein the error δ_{pj} of each of the units u_{H1} to u_{Hy} of the intermediate layer L_H is given by the following formula (22)

$$\delta_{Hj} = O_{Hj}(1 - O_{Hj}) \sum_k \delta_{ok} W_{kj} \quad (22)$$

[0042] Then, in step 4, the learning variable β_j of the coefficient W_{ji} of coupling strength from the i'th one to the j'th one of the units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz} is computed by the following formula (23)

$$\beta_j = \frac{1}{\sum_i Opi^2 + 1} \quad (23)$$

in which the learning variable β_j is represented by the reciprocal of the square sum of the input values added to by 1 as a threshold value.

[0043] Then, in step 5, using the learning variable β_j computed in step 4, the variant Δw_{ij} of the coupling coefficient W_{ji} from the i'th one to the j'th one of the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} is computed in accordance with the following formula (24):

$$\Delta W_{ji(n)} = \eta \cdot \beta(\delta_{pj} O_{pi}) \quad (24)$$

In the formula, η stands for a learning constant.

[0044] Then, in step 5, the total sum LMS of the square errors on the units with respect to the teacher signal is computed in accordance with the formula (25):

$$LMS = \sum_{p=1} \sum_{i=1} (t_{pi} - O_{pi}) \quad (25)$$

[0045] Then, in step 6, it is decided whether the processing of the steps 1 through 5 has been performed on the R-number of input signal patterns p_{xr}. If the result of decision at step 6 is NO, the section 40 reverts to step 1. When the result of decision at step 6 is YES, that is, when all of the variants ΔW_{ji} of the coupling coefficient w_{ji} between the units u_{O1} to u_{O2}, u_{H1} to u_y and u_{I1} to U_{Ix} are computed for the input signal patterns p_{xr}, the section 40 proceeds to step 7 to execute decision of the converging condition for the output value O_{oj} obtained at the output layer L_O on the basis of the total sum LMS of square errors between the output value O_{oj} and the desired output value t_{pj} afforded as the teacher signal.

[0046] In the decision step 7, it is decided whether the output value O_{oj} obtained at the output layer L_O of the signal processing section 30 is closest to the desired output value t_{pj} afforded as the teacher signal. When the result of decision at step 7 is YES, that is, when the total sum LMS of the square errors is sufficiently small and the output value O_{oj} is closest to the desired output value t_{pj}, the learning processing is terminated. If the result of decision at step 7 is NO, the section 40 proceeds to computing at step 8.

[0047] In this computing step 8, the coupling coefficient W_{ji} between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} is modified, on the basis of the variant W_{ji} of the coupling coefficient W_{ji} computed at step 5, in accordance with the following formula (26)

$$\Delta W_{ji(n)} = \Delta W_{ji(n)} + \alpha \Delta W_{ji(n-1)} \quad (26)$$

and the following formula (27)

$$W_{ji(n+1)} = W_{ji(n)} + \Delta_{Wji(n)} \quad (27)$$

[0048] After the computing step 8, the section 40 reverts to step 1 to execute the operation of steps 1 to 6.

[0049] Thus the section 40 executes the operations of the steps 1 to 8 repeatedly and, when the total sum LMS of the square errors between the desired output value t_{pj} and the actual output value O_{oj} becomes sufficiently small and the output value O_{oj} obtained at the output value L_O of the signal processing section 30 is closest to the desired output value t_{pj} afforded as the teacher signal, terminates the processing of learning by the decision at step 7.

[0050] In this manner, in the illustrative example signal processing system, the learning as to the coupling coefficient W_{ji} between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} of the signal processing section 30 constituting the recurrent network inclusive of the above mentioned loop LP and the feedback FB is executed by the learning processing section 40 on the basis of the desired output value t_{pj} afforded as the teacher signal. Hence, the features of the sequential time-base input signal pattern p_{xr}, such as audio signals, fluctuating along the time axis, may also be extracted reliably by the learning processing by the learning processing section 40. Thus, by setting the coupling state between the units u_{O1} to u_{Oz}, u_{H1} to u_{Hy} and u_{I1} to u_{Ix} of the signal processing section 30 by the coupling coefficient W_{ji}, obtained as the result of learning by the learning processing section 40, the time-series input signal pattern p_{xr} can be subjected to desired signal processing by the signal processing section 30.

[0051] Moreover, in the illustrative example system, the learning constant η is normalized by the learning constant β indicated as the reciprocal of the square sum of the input values at the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, and the learning processing as to the coupling coefficient W_{ji} is performed at the dynamically changing learning rate, as a function of the input value O_{pi}, so that learning can be performed stably and expeditiously with a small number of times of learning.

[0052] In this manner, in the illustrative example signal processing system, signal processing for input signals is performed at the signal processing section 30 in which the recurrent network inclusive of the loop LP and the feedback FB is constituted by the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} of the intermediate layer L_H and the output layer L_O each provided with delay means. In the learning processing section 40, the learning as to the coupling state of the recurrent network by the units u_{H1} to u_{Hy} and u_{O1} to u_{Oz} constituting the signal processing section 30 is executed on the basis of the teacher signal. Thus the features of the sequential time-base patterns, fluctuating along the time axis, such as audio signals, can be extracted by the above mentioned learning processing section to subject the signal processing section to the desired signal processing.

[0053] A preferred illustrative embodiment learning processing system according to the present invention will be hereinafter explained.

[0054] The basic construction of the learning processing system according to the present invention is shown in Fig. 5. As shown therein, the system includes a signal processing section 50 constituted by a neural network of a three-layered structure including at least an input layer L_I, an intermediate layer L_H and an output layer L_O, each made up of plural units performing a signal processing corresponding to one of a neuron, and a learning processing section 60 subjecting the learning processing to the signal processing consisting in sequentially repeatedly computing the coefficient W_{ji} of coupling strength between the above units from the output layer L_O towards the input layer L_I on the basis of the error data δ_{pj} between the output value of the output layer L_O and the desired output value O_{pj} afforded as the teacher signal t_{pj}, for the input signal patterns p entered into the input layer L_I of the signal processing section 50, and learning the coupling coefficient W_{ji} in accordance with the back-propagation learning rule.

[0055] The learning processing section 60 executes the learning processing of the coupling coefficient W_{ji} as it causes the number of the units of the intermediate layer L_H of the signal processing section 50 to be increased, and thus the section 60 has the control function of causing the number of units of the intermediate layer L_H to be increased in the course of learning processing of the coupling coefficient W_{ji}. The learning processing section 60 subjects the signal processing section 50 having the input layer L_I, an intermediate layer L_H and an output layer L_O made up of arbitrary numbers x, y and z of units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, each corresponding to a neuron, respectively, as shown in Fig. 6A, to learning processing as to the coupling coefficient W_{ji}, while the section 60 causes the number of the unit L_H to be increased sequentially from y to (y+m), as shown in Fig. 6B.

[0056] It is noted that the control operation of increasing the number of the units of the intermediate layer L_H may be performed periodically in the course of learning processing of the coupling coefficient W_{ji}, or each time the occurrence of the above mentioned local minimum state is sensed.

[0057] The above mentioned learning processing section 60, having the control function of increasing the number of the units of the intermediate layer L_H in the course of learning processing of the coupling coefficient W_{ji}, subjects the signal processing section 50 formed by a neural network of a three-layer structure including the input layer L_I, intermediate layer L_H and the output layer L_O to the learning processing of the coupling coefficient W_{ji}, as it causes the number of units of the intermediate layer L_H to be increased. Thus, even on occurrence of the local minimum state in the course of learning of the coupling coefficient W_{ji}, the section 50 is able to increase the number of units of the intermediate layer L_H to exit from such local minimum state to effect rapid and reliable convergence into the optimum global minimum state.

[0058] Tests were conducted repeatedly, in each of which the learning processing section 50 having the control

function of increasing the number of units of the intermediate layer in the course of learning of the coupling coefficient W_{ji} causes the signal processing section 60 constituting the recurrent network including the feedback FB and the loop LP in the illustrative example, signal processing system of Figs. 2-4 to undergo the process of learning the coefficient W_{ji}, with the number of the units of the input layer L_I of 8(x=8), that of the output layer L_O of 3(z=3), the number of the delay means of each layer of 2 and with the input signal pattern p during learning operation, using 21 time-space patterns of 1=8x7, and the processing algorithm shown in the flow chart of Fig. 7, with the learning being started at the number of the units of the intermediate layer L_H of 3(y=3) and with the number of the units of the intermediate layer L_H being increased during the learning process. By increasing the number of the units of the intermediate layer L_H three to five times, the test results were obtained in which the convergence to the optimum global minimum state were realized without going into the local minimum state.

[0059] Fig. 8 shows, as an example of the above tests, the test results in which learning processing of converging into the optimum minimum state could be achieved by adding the units of the intermediate layer L_H at the timing shown by the arrow mark in the figure and by increasing the number of the intermediate layer L_H from three to six. The ordinate in Fig. 8 stands for the total sum LMS of the quadratic errors and the abscissa stands for the number of times of the learning processing operations.

[0060] The processing algorithm shown in the flow chart of Fig. 7 is explained.

[0061] In this processing algorithm, in step 1, the variable K indicating the number of times of the processing for detecting the local minimum state is initialized to "0", while the first variable Lms for deciding the converging condition of the learning processing is also initialized to 10000000000.

[0062] Then, in step 2, the variable n indicating the number of times of learning of the overall learning pattern, that is, the l-number of the input signal patterns p, is initialized. The program then proceeds to step 3 to execute the learning processing of the l-number of the input signal patterns p.

[0063] Then, in step 4, decision is made of the variable n indicating the number of times of learning. Unless n=3, the program proceeds to step 5 to add one to n ($n \rightarrow n+1$), and then reverts to step 3 to repeat the learning processing. When n=3, the program proceeds to step 6.

[0064] In step 6, after the value of the first variable Lms is maintained as the value of the second variable Lms (-1) for deciding the converging condition of the learning processing, the total sum of the square errors between the output signal and the teacher signal in each unit is computed in accordance with the formula (28), this value being then used as the new value for the first variable Lms, such that

$$Lms = \sum_{p=1}^{l} \sum_{i=1}^{m} (t_{pi} - O_{pi})^2 \qquad (28)$$

[0065] Then, in step 7, the first variable Lms for deciding the converging condition of the learning processing is compared with the second variable Lms(-1). If the value of the first variable Lms is lesser than that of the second variable Lms(-1), the program proceeds to step 8 to decide whether or not the variable k indicating the number of times of the processing operations for detecting the local minimum state is equal to 0.

[0066] If, in step 8, the variable k is 0, the program reverts directly to step 2. If the variable k is not 0, setting of $k \rightarrow k-1$ is made in step 9. The program then reverts to step 2 to initialize n to 0(n=0) to execute the learning processing of the l-number of the input signal patterns p in step 3.

[0067] If, in step 7, the value of the first variable Lms is larger than that of the second variable Lms(-1), the program proceeds to step 10 to set the value of k indicating the number of times of the processing operations for detecting the local minimum state ($k \rightarrow k + 1$). Then, in step 11, it is decided whether or not the value of k is 2.

[0068] If, in step 11, the value of the variable k is not 2, the program reverts directly to step 2. If the variable k is 2, it is decided that the local minimum state is prevailing. Thus, in step 12, control is made of increasing the number of the units of the intermediate layer L_H. Then, in step 13, setting of k=0 is made. The program then reverts to step 2 for setting of n=0 and then proceeds to step 3 to execute the learning processing of the above mentioned l-number of the input signal pattems p.

[0069] Test on the learning processing was conducted of the signal processing section 50 of the above described illustrative example signal processing system of Figs. 2-4 constituting the recurrent network including the feedback loop FB and the loop LP shown in Fig. 3, with the number of the units of the intermediate layer L_H being set to six (y=6). The test results have revealed that the learning processing need be repeated an extremely large number of times with considerable time expenditure until the convergence to the optimum minimum state was achieved, and that the local minimum state prevailed for three out of eight learning processing tests without convergence to the optimum global minimum state.

[0070] Fig. 9 shows, by way of an example, the results of the learning processing tests in which the local minimum

state was reached.

[0071] In this figure, the ordinate stands for the total sum LMS of the square errors and the abscissa stands for the number of times of the learning processing operations.

[0072] Also the tests on the learning processing was conducted 30 times on the signal processing section 50 of the above described illustrative example signal processing system constituting the recurrent network including the feedback loop FB and the loop LP shown in Fig. 3, with the number of the units of the intermediate layer L_H being set to three (y=3). It was found that, as shown for example in Fig. 10, the local minimum state was reached in all of the tests on learning processing without convergence to the optimum global minimum state.

[0073] In Fig. 10, the ordinate stands for the total sum LMS of the square errors and the abscissa stands for the number of times of the learning processing operations.

[0074] From the foregoing it is seen that the present invention provides a learning processing system in which the learning processing of the coefficient of coupling strength is performed, while the number of the units of the intermediate layer is increased by the learning processing section, whereby the convergence to the optimum global minimum state is achieved promptly and reliably to achieve the stable learning processing to avoid the local minimum state in the learning processing process conforming to the back-propagation learning rule.

Claims

1. A learning processing system (50, 60) comprising:

   a signal processing section (50) composed of a multi-layer neural network having an input layer (L_I), a bidden layer (L_H) and an output layer (L_O), the layers being made up of units u_{I1} to u_{Ix}, u_{H1} to u_{Hy} and u_{O1} to u_{Oz}, respectively, each unit corresponding to a neuron; and
   a learning processing section (60) executing a learning process using a back-propagation learning algorithm, the process consisting in sequentially modifying, from the output layer towards the input layer, the coupling coefficients W_{ji} of all units j in the hidden and in the output layer by a variant ΔW_{ji} so as to minimize the total sum of square errors between the actual output O_{pj} of unit j in the output layer (L_O) produced from an input signal pattern (p) and the desirable output value t_{pj} (teacher signal) for said unit j in the output layer (L_O), whereby W_{ji} is the weight for the signal from the ith to the jth unit,
   the learning processing section being fed with a desired output value t_{pj} as a teacher signal for the output value O_{pj} of the unit j in the output layer (L_O) for the input patterns p entered into the input layer (L_I),
   the learning processing section (60) computing the error value for each unit in the output layer and in the bidden layer,
   said learning process being executed repeatedly until the total sum (E) of the square error between the desired output afforded as the teacher signal and the output signal becomes sufficiently small;

   characterised in that the learning processing section (60) comprises control means for increasing the number of units in the bidden layer (L_H), during the repeated execution of said learning process, either periodically or when a local minimum state of the signal processing system has been detected, said learning processing section (GO) being adapted in subsequent repeated executions of the learning process to perform learning processing of the coefficients W_{ji} of coupling strength in respect of the increased number of units in the hidden layer.

2. The learning processing system of claim 1, wherein the control means of the learning processing section (60) is adapted to detect a local minimum state by comparing successive values of a first variable, Lms, where

$$Lms = \sum_{p=1}^{l} \sum_{i=1}^{m} (t_{pi} - O_{pi})^2 .$$

Patentansprüche

1. Lernverarbeitungssystem (50, 60)

   mit einem Signalverarbeitungsabschnitt (50), der aus einem neuronalen Mehrschichten-Netzwerk mit einer

Eingangsschicht (L_I), einer verborgenen Schicht (L_H) und einer AusgangsSChiCht (L_O) zusammengesetzt ist, wobei die Schichten aus Einheiten u_{I1} bis u_{Ix}, u_{H1} bis u_{Hy} bzw. u_{O1} bis u_{Oz} bestehen und jede Einheit einem Neuron entspricht,
und mit einem Lernverarbeitungsabschnitt (60), der einen Lernprozeß unter Verwendung eines sich rückwärts ausbreitenden Algorithmus ausführt, wobei der Prozeß darin besteht, daß von der Ausgangsschicht in Richtung zu der Eingangsschicht die Kopplungskoeffizienten W_{ji} aller Einheiten j in der verborgenen Schicht und in der Ausgangsschicht durch eine Variante ΔW_{ji} so modifiziert werden, daß die Gesamtsumme der quadratischen Fehler zwischen dem durch ein Eingangssignalmuster (p) erzeugten tatsächlichen Ausgangswert O_{pj} der Einheit in der Ausgangsschicht L_O und dem gewünschten Ausgangswert t_{pj} (Lehrersignal) für diese Einheit j in der Ausgangsschicht (L_O) minimiert wird, wobei W_{ji} das Gewicht für das Signal aus der i-ten Einheit zu der j-ten Einheit ist,
wobei dem Lernverarbeitungsabschnitt ein gewünschter Ausgangswert t_{pj} als Lehrersignal für den Ausgangswert O_{pj} der Einheit j in der Ausgangsschicht (L_O) für die in die Eingangsschicht (L_I) eingegebenen Eingangsmuster p zugeführt wird,
wobei der Lernverarbeitungsabschnitt (60) den Fehlerwert für jede Einheit in der Ausgangsschicht und in der verborgenen Schicht berechnet,
und wobei der Lernprozeß wiederholt durchgeführt wird, bis die Gesamtsumme (E) des quadratischen Fehlers zwischen dem als Lehrersignal angebotenen gewünschten Signal und dem Ausgangssignal hinreichend klein wird,

dadurch gekennzeichnet,

daß der Lernverarbeitungsabschnitt (60) eine Steuereinrichtung enthält, um die Zahl der Einheiten in der verborgenen Schicht (L_H) während der wiederholten Durchführung des Lernprozesses entweder periodisch oder dann zu vergrößern, wenn ein lokaler Minimalzustand des Signalverarbeitungssystems detektiert wurde, wobei der Lernverarbeitungsabschnitt (60) so ausgebildet ist, daß er in den nachfolgenden wiederholten Durchführungen des Lernprozesses die Lernverarbeitung der Kopplungskoeffizienten W_{ji} bezüglich der vergrößerten Zahl von Einheiten in der verborgenen Schicht durchführt.

2. Lernverarbeitungssystem nach Anspruch 1, bei dem die Steuereinrichtung des Lern-Verarbeitungsabschnitts (60) den lokalen Minimalzustand durch Vergleichen von aufeinanderfolgenden Werten einer ersten Variablen Lms detektieren kann, wobei

$$Lms = \sum_{p=1}^{l} \sum_{i=1}^{m} (t_{pi} - O_{pi})^2.$$

Revendications

1. Système de traitement d'apprentissage (50, 60) comprenant :

une section de traitement de signaux (50) constituée d'un réseau neuronal multicouche ayant une couche d'entrée (L_I), une couche intermédiaire (L_H) et une couche de sortie (L_O), les couches étant respectivement formées d'unités u_{I1} à u_{Ix}, u_{H1} à u_{Hy}, u_{O1} à u_{Oz}, chaque unité correspondant à un neurone ;
une section de traitement d'apprentissage (60) qui exécute un traitement d'apprentissage utilisant un algorithme d'apprentissage de rétropropagation, ce traitement consistant à modifier séquentiellement, de la couche de sortie vers la couche d'entrée, les coefficients de couplage W_{ji} de toutes les unités j de la couche intermédiaire et de la couche de sortie à l'aide d'un coefficient de variation ΔW_{ji} de façon à minimiser la somme totale des erreurs quadratiques existant entre la valeur de sortie réelle O_{pj} de l'unité j de la couche de sortie (L_O) produite à partir d'une forme de signal d'entrée (p) et la valeur de sortie souhaitable t_{pj} (signal d'enseignement) pour ladite unité j de la couche de sortie (L_O), si bien que W_{ji} est le poids du signal de la i$^{\text{ème}}$ à la j$^{\text{ème}}$ unité,
la section de traitement d'apprentissage recevant une valeur de sortie voulue t_{pj} au titre d'un signal d'enseignement, pour la valeur de sortie O_{pj} de l'unité j de la couche de sortie (L_O), pour les formes d'entrée p introduites dans la couche d'entrée (L_I),
la section de traitement d'apprentissage (60) calculant la valeur d'erreur pour chaque unité de la couche de sortie et de la couche intermédiaire,

ledit traitement d'apprentissage étant exécuté de façon répétée jusqu'à ce que la somme totale (E) des erreurs quadratiques existant entre le signal de sortie souhaité, qui est fourni au titre du signal d'enseignement, et le signal de sortie soit devenue suffisamment petite,

caractérisé en ce que la section de traitement d'apprentissage (60) comprend un moyen de commande servant à augmenter le nombre des unités de la couche intermédiaire (L_H), pendant l'exécution répétée dudit traitement d'apprentissage, ou bien périodiquement, ou bien lorsqu'un état de minimum local du système de traitement de signaux a été détecté, ladite section de traitement d'apprentissage (60) étant conçue pour, lors d'exécutions répétées suivantes du traitement d'apprentissage, effectuer le traitement d'apprentissage des coefficients W_{ji} d'intensité de couplage relativement au nombre augmenté d'unités de la couche intermédiaire.

2. Système de traitement d'apprentissage selon la revendication 1, où le moyen de commande de la section de traitement d'apprentisssage (60) est conçu pour détecter un état de minimum local par comparaisons de valeurs successives d'une première variable, L_{ms}, où :

$$Lms = \sum_{p=1}^{l} \sum_{i=1}^{m} (t_{pi} - O_{pi})^2$$

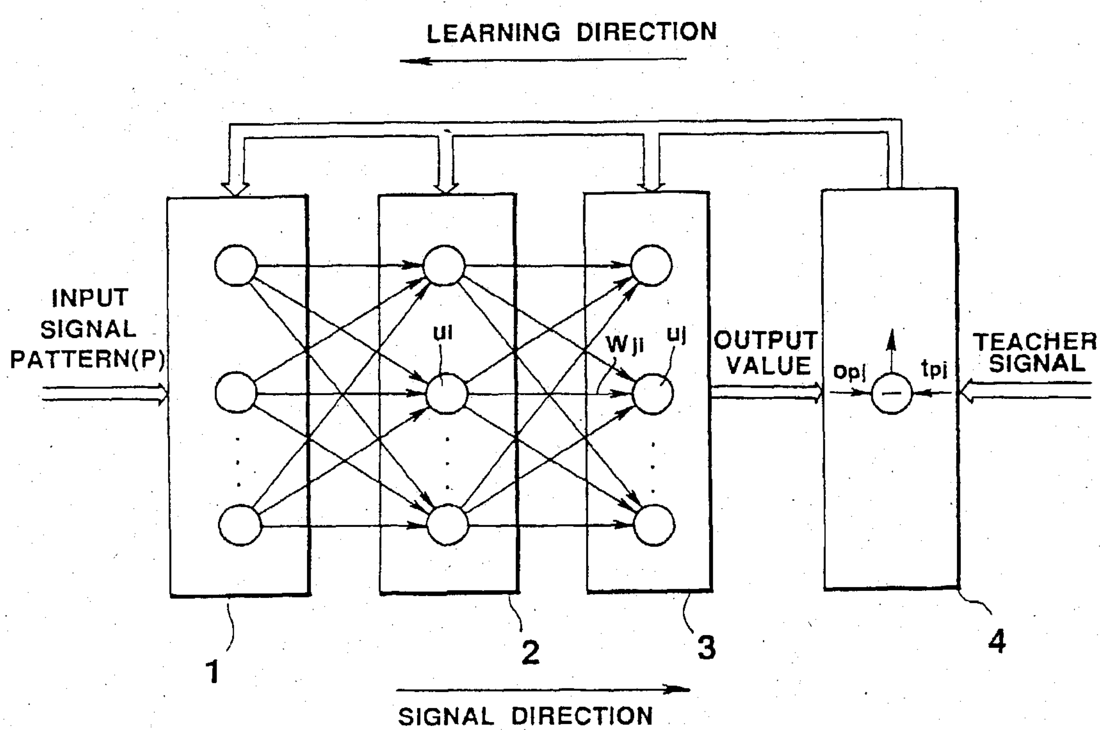
LEARNING DIRECTION
INPUT SIGNAL PATTERN(P)
ui
Wji
uj
OUTPUT VALUE
opj
tpj
TEACHER SIGNAL
1
2
3
4
SIGNAL DIRECTION

FIG. 1

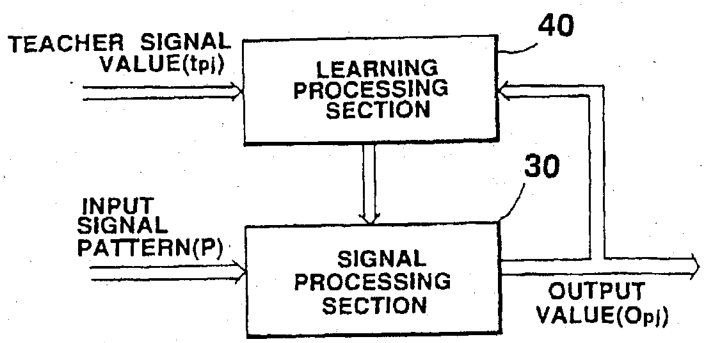
TEACHER SIGNAL VALUE(tpj)
40
LEARNING PROCESSING SECTION
30
INPUT SIGNAL PATTERN(P)
SIGNAL PROCESSING SECTION
OUTPUT VALUE(Opj)

FIG.2

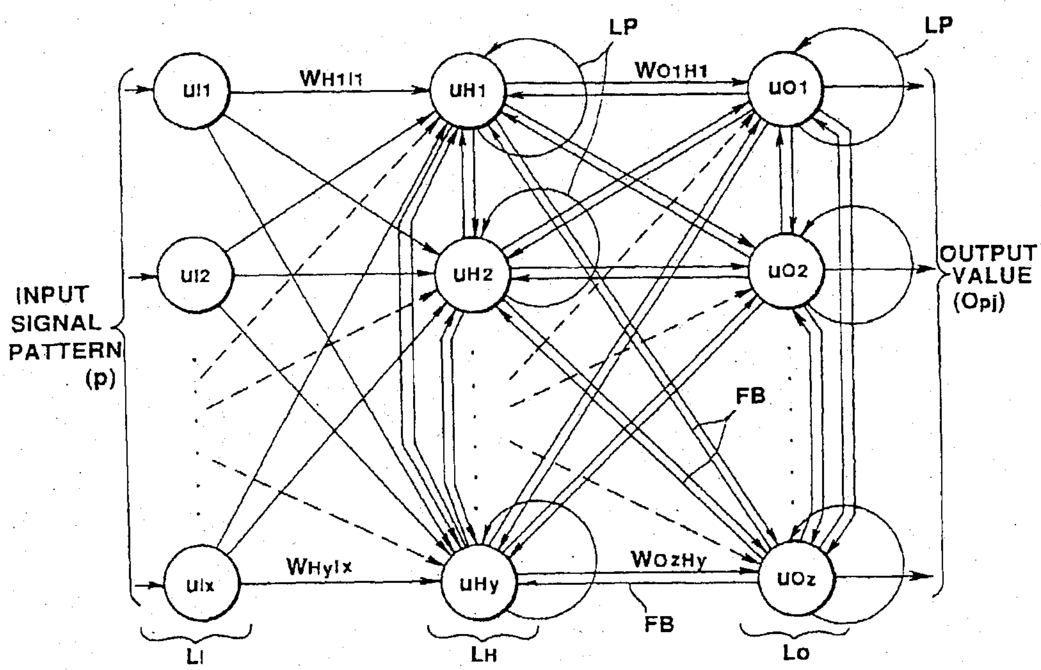
LP
LP
uI1
WH1I1
uH1
WO1H1
uO1
uI2
uH2
uO2
INPUT SIGNAL PATTERN (p)
OUTPUT VALUE (Opj)
FB
uIx
WHyIx
uHy
WOzHy
uOz
FB
LI
LH
LO

FIG.3

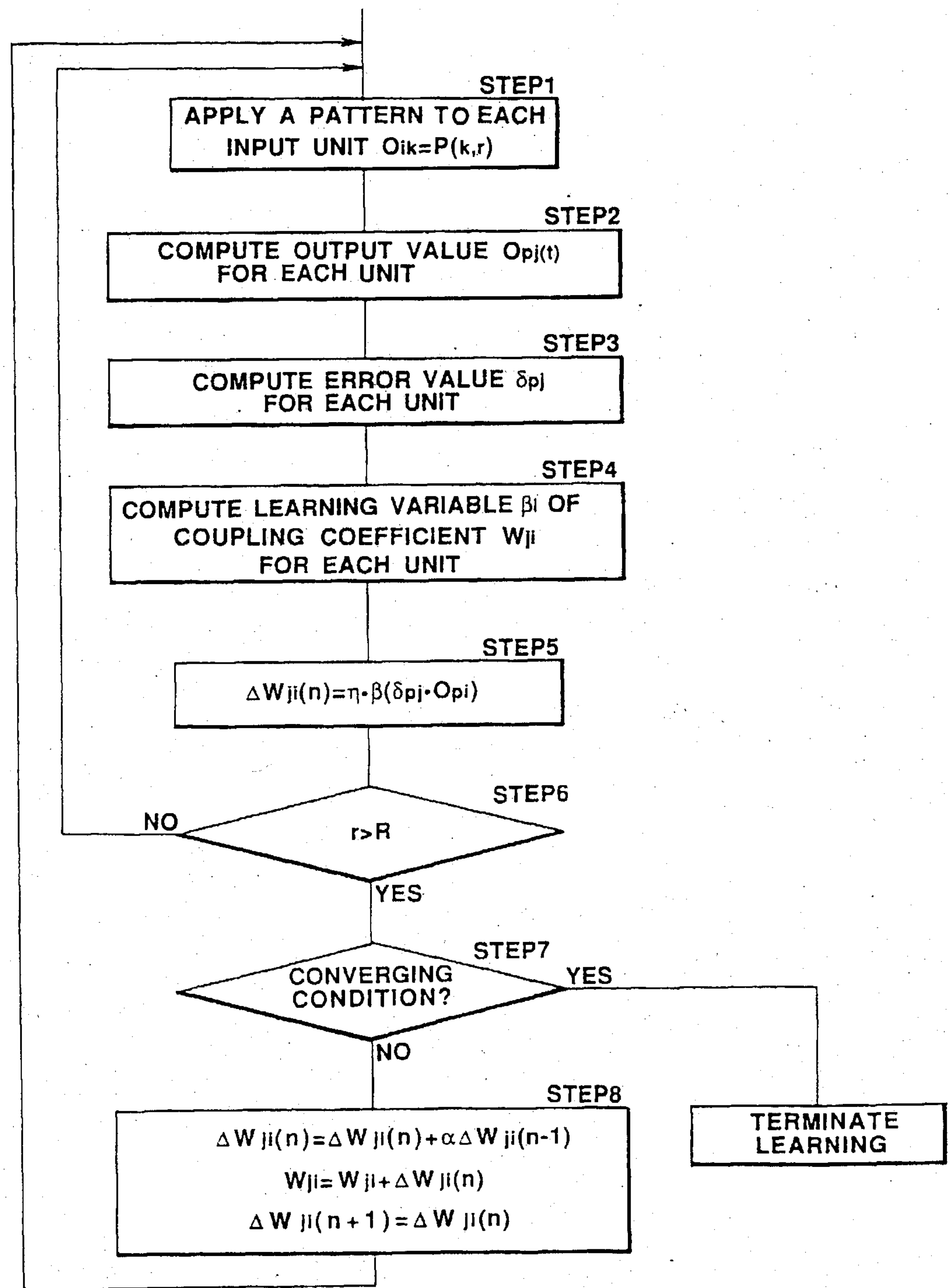
STEP1
APPLY A PATTERN TO EACH INPUT UNIT $O_{ik}=P(k,r)$
STEP2
COMPUTE OUTPUT VALUE $O_{pj}(t)$ FOR EACH UNIT
STEP3
COMPUTE ERROR VALUE δ_{pj} FOR EACH UNIT
STEP4
COMPUTE LEARNING VARIABLE β_i OF COUPLING COEFFICIENT W_{ji} FOR EACH UNIT
STEP5
$\Delta W_{ji}(n)=\eta\cdot\beta(\delta_{pj}\cdot O_{pi})$
STEP6
$r>R$
NO
YES
STEP7
CONVERGING CONDITION?
YES
NO
STEP8
$\Delta W_{ji}(n)=\Delta W_{ji}(n)+\alpha\Delta W_{ji}(n-1)$
$W_{ji}=W_{ji}+\Delta W_{ji}(n)$
$\Delta W_{ji}(n+1)=\Delta W_{ji}(n)$
TERMINATE LEARNING

FIG.4

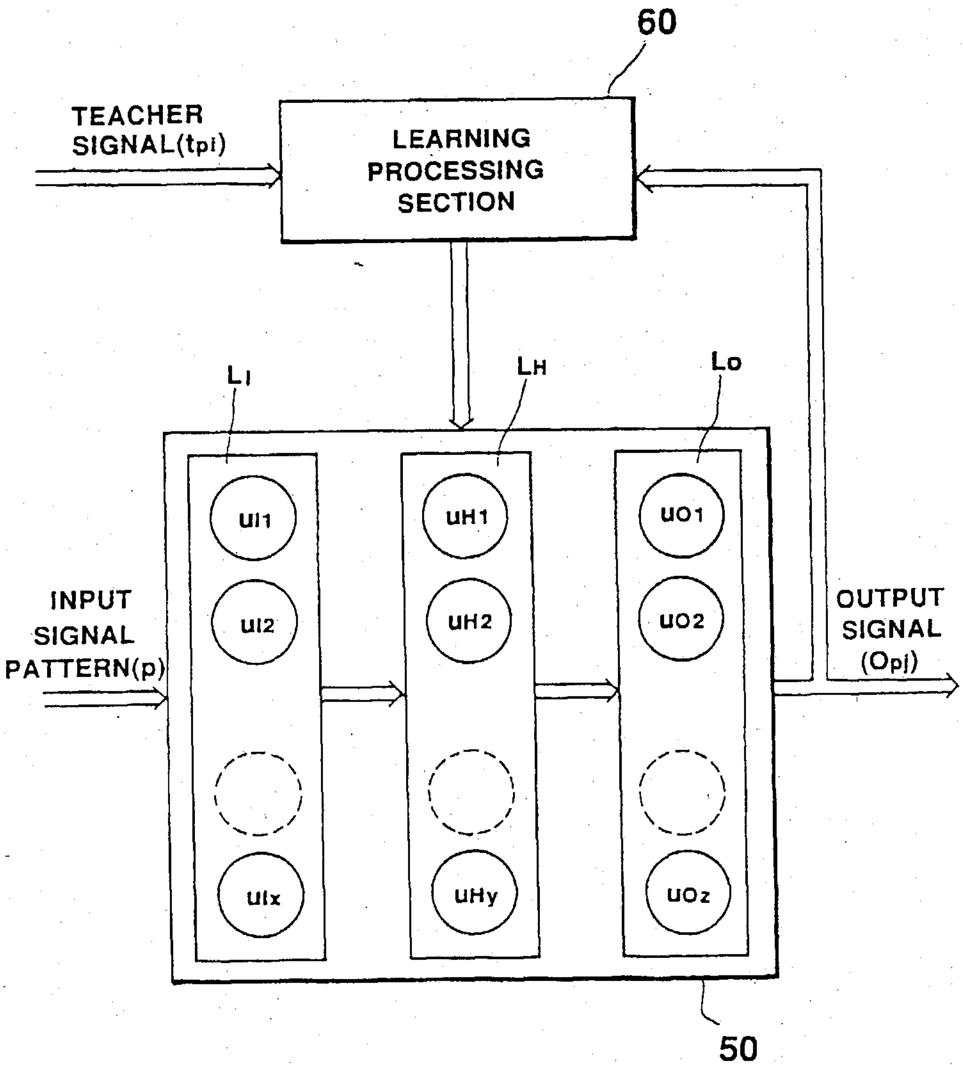
60
TEACHER SIGNAL(tpi)
LEARNING PROCESSING SECTION
LI
LH
Lo
uI1
uI2
uIx
uH1
uH2
uHy
uO1
uO2
uOz
INPUT SIGNAL PATTERN(p)
OUTPUT SIGNAL (Opj)
50

FIG.5

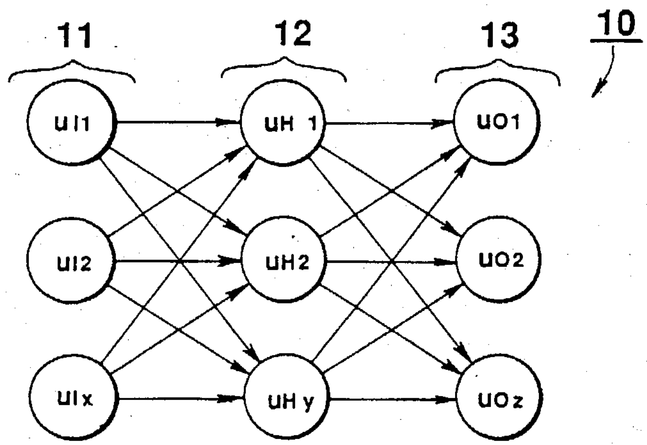
11
12
13
10
uI1
uH 1
uO1
uI2
uH2
uO2
uIx
uH y
uOz

FIG.6(A)

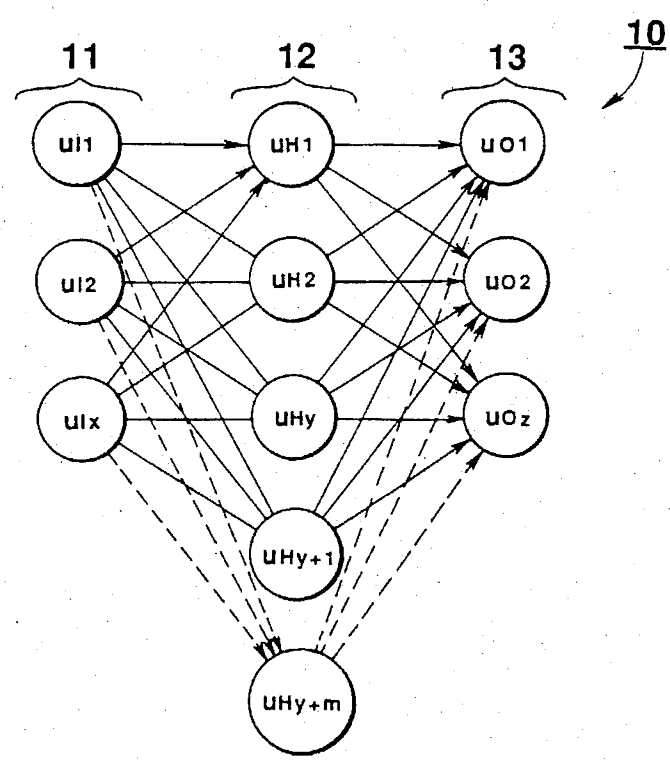
11
12
13
10
uI1
uH1
uO1
uI2
uH2
uO2
uIx
uHy
uOz
uHy+1
uHy+m

FIG.6(B)

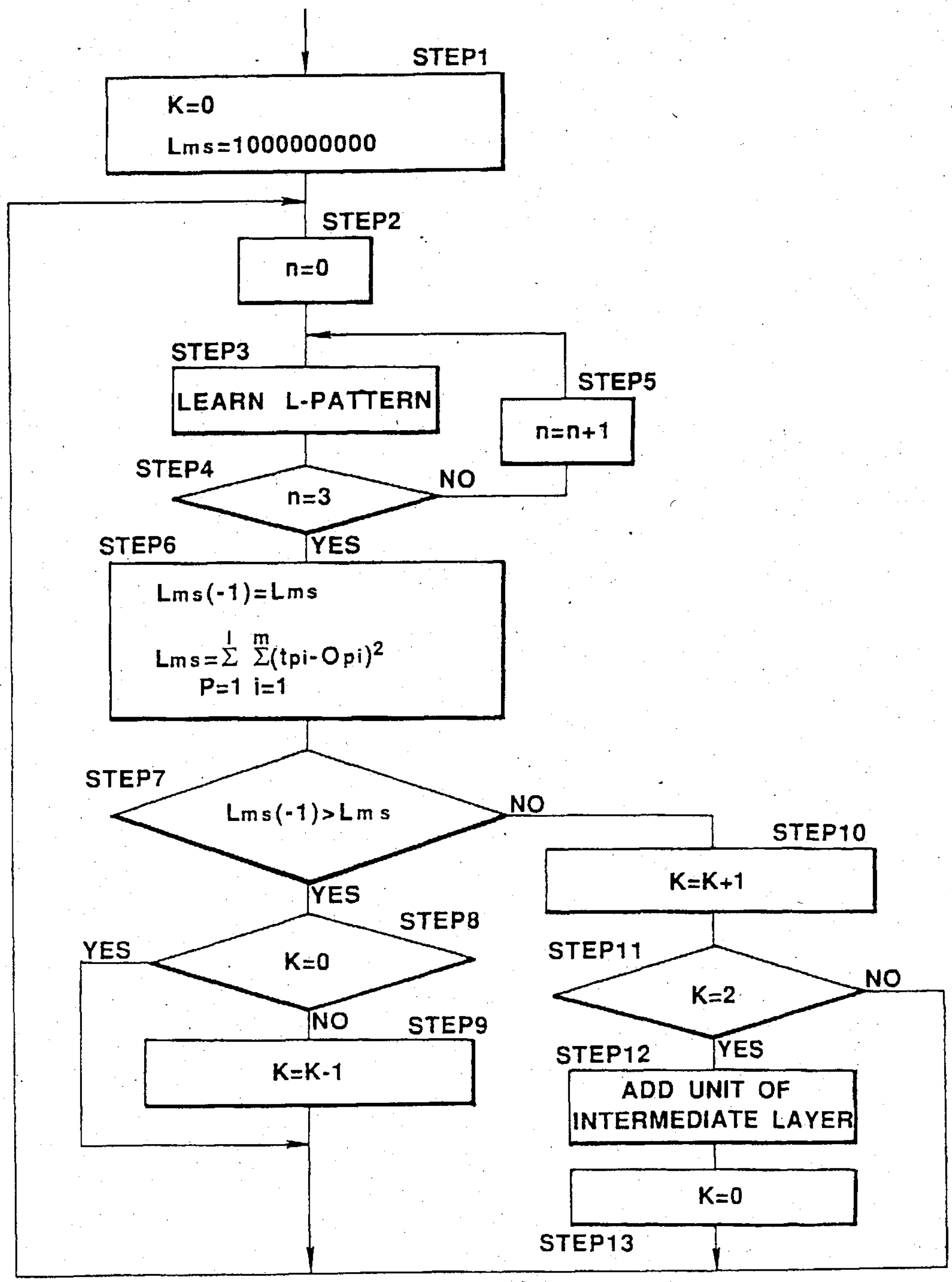
STEP1
K=0
Lms=1000000000
STEP2
n=0
STEP3
LEARN L-PATTERN
STEP5
n=n+1
STEP4
n=3
NO
YES
STEP6
Lms(-1)=Lms
$Lms=\sum_{P=1}^{l}\sum_{i=1}^{m}(tpi-Opi)^2$
STEP7
Lms(-1)>Lms
NO
YES
STEP10
K=K+1
STEP8
K=0
YES
NO
STEP9
K=K-1
STEP11
K=2
NO
YES
STEP12
ADD UNIT OF
INTERMEDIATE LAYER
K=0
STEP13

FIG. 7

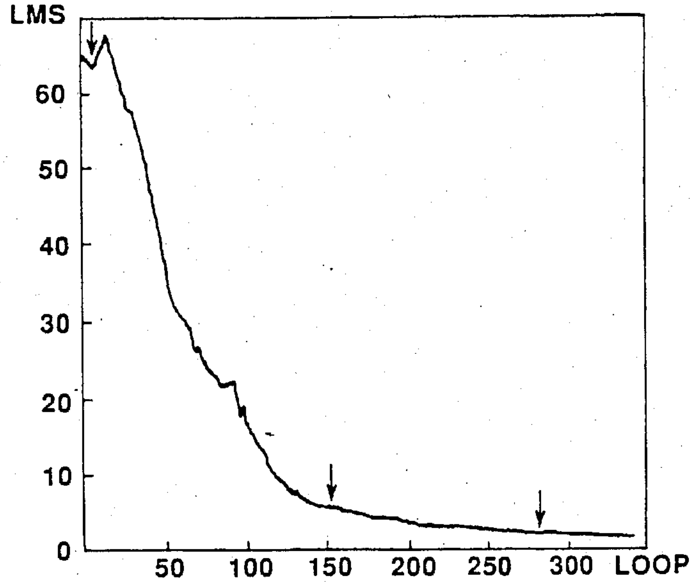
LMS
60
50
40
30
20
10
0
50
100
150
200
250
300
LOOP

FIG. 8

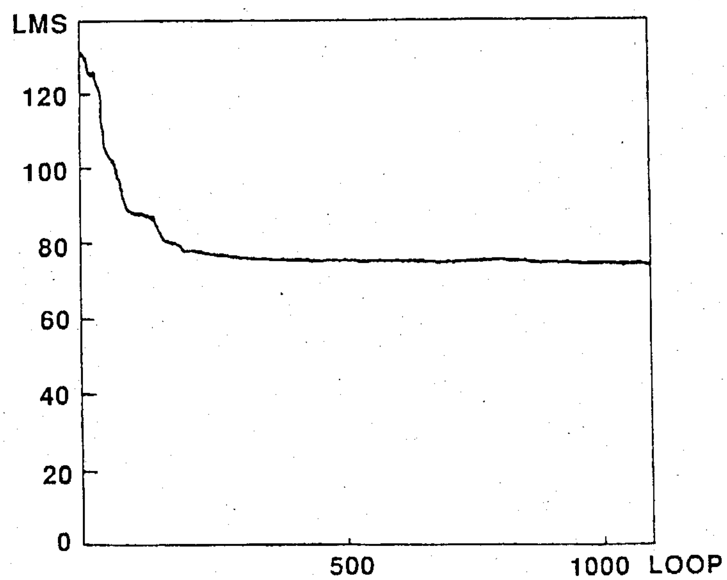
LMS
120
100
80
60
40
20
0
500
1000
LOOP

FIG. 9

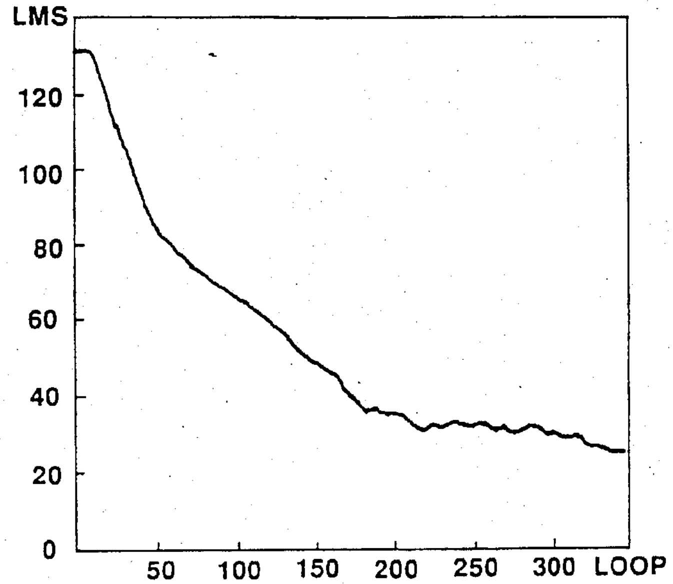

FIG. 10